On the Variation of the Spectra of Matrices

L. Elsner
Fakultät für Mathematik
Universität Bielefeld
Postfach 8640
4800 Bielefeld, Federal Republic of Germany

Submitted by S. Friedland

ABSTRACT

Several "distances" between the spectra of two matrices are discussed and compared. Optimal bounds are given, which enable us to reduce certain bounds on the eigenvalue variation of matrices by a factor of about two. The results of Bhatia and Mukherjea and of Bhatia and Friedland on the eigenvalue variation are derived in an elementary way using results of Henrici on the spectral variation.

INTRODUCTION

It is well known that the eigenvalues of an $n\times n$ matrix A depend continuously on the elements of A. In many applications, e.g. inverse eigenvalue problems, more specific information is required. For the general case, quantitative results on the change of the spectrum have been obtained by Ostrowski [11], by Henrici [8], and recently by Bhatia and Mukherjea [3] and by Bhatia and Friedland [5]. It is the aim of this paper to develop their results.

The two main results are the following.

In Theorem 1 we derive a comparison between two measures for the distance between the spectra of two matrices, which is sharp. It allows us to reduce the bounds on the eigenvalue variation of two matrices in the general case, which are mentioned above, by a factor of about 2.

Then we give a new derivation and slight improvement of the results in [3] and [5], based on Henrici's result in [8]. Indeed, they can be obtained by

*This research was performed while the author was visiting the University of Calgary in 1980–1981. He is grateful for hospitality and financial support. This note has also appeared as Research Paper No. 482 (April 1981) of the Department of Mathematics and Statistics, the University of Calgary.

52 Vanderbilt Ave., New York, NY 10017 0024-3795/82/060127-12$02.75

elementary inequality manipulations with real functions. Together with the fact that Henrici's bound is proved using simple norm estimates of the resolvent, this shows that a short and elementary derivation is possible, avoiding the use of the characteristic equation.

1. NOTATION AND BASIC RESULTS

For two complex $n \times n$ matrices A and B with spectra $\sigma(A) = \{\lambda_1, \dots, \lambda_n\}$ and $\sigma(B) = \{\mu_1, \dots, \mu_n\}$, where the eigenvalues are counted according to their algebraic multiplicities, we introduce the following "distances" between the spectra:

$$S_A(B) = \max_i \min_j |\lambda_j - \mu_i|, \tag{1.1}$$

the spectral variation of B with respect to A (see [7]);

$$h_A(B) = \max_{0 \leqslant t \leqslant 1} S_A((1-t)A + tB); \tag{1.2}$$

and

$$v(A, B) = \min_\pi \max_i |\lambda_i \mu_{\pi(i)}|, \tag{1.3}$$

the eigenvalue variation of A and B. Here the minimum is taken over all permutations π of $\{1, 2, \dots, n\}$. We shall use the Euclidean matrix norm

$$\| A \|_E = \left(\sum_{i,k} |\alpha_{ik}|^2 \right)^{1/2} \tag{1.4a}$$

and the spectral norm

$$\| A \|_2 = \rho(A^H A)^{1/2}. \tag{1.4b}$$

Here $\rho(A)$ denotes the spectral radius of A.

The main result connecting the different "distances" defined above is the following.

THEOREM 1. *For two complex $n \times n$ matrices A, B,*

$$v(A, B) \leqslant (2n-1) h_A(B) \tag{1.5}$$

and

$$v(A,B) \leqslant a_n \max\{h_A(B), h_B(A)\}, \tag{1.6}$$

where

$$a_n = \begin{cases} n & \textit{if } n \textit{ is odd}, \\ n-1 & \textit{if } n \textit{ is even}. \end{cases} \tag{1.7}$$

Moreover, the constants $2n-1$ *and* a_n *cannot be improved.*

Proof. (1.5) may be proved adapting the argument used by Ostrowski in [11, p. 276 ff.]. This adaptation is not new: see the remarks in [8, 4.4, p. 35] and the remark referred to in that remark. For completeness let us indicate this proof.

We may interpret $h_A(B)$ geometrically by saying that the spectra of $(1-t)A+tB$, $t \in [0,1]$, are contained in

$$\tilde{K}_A = \bigcup_{i=1}^{n} C(\lambda_i, h_A(B)),$$

where $C(\lambda; r) = \{z \in C: |z-\lambda| \leqslant r\}$. As the eigenvalues of $(1-t)A+tB$ depend continuously on t, each connected component of $\tilde{K}_A$ contains as many eigenvalues of A as eigenvalues of B. Matching eigenvalues in the same connected component yields (1.5).

Turning to the proof of (1.6), denote $\max(h_A(B), h_B(A))$ by δ. We may interpret this geometrically by saying that the spectra of $(1-t)A+tB$ are contained in

$$K_A = \bigcup C(\lambda_i, \delta) \quad \text{and} \quad K_B = \bigcup C(\mu_i, \delta).$$

As shown above, each connected component of K_A and of K_B contains as many eigenvalues of A as of B. In [7] the following theorem is proved by graph-theoretical means, using Hall's theorem.

THEOREM 2. *Let* $\lambda_1 \ldots \lambda_n$, $\mu_1, \ldots, \mu_n$ *be* $2n$ *points in the complex plane, such that each connected component of* $\bigcup_{i=1}^{n} C(\lambda_i, 1)$ *containing* p *circles contains exactly* p *of the numbers* $\mu_1 \ldots \mu_n$ *and each connected component of* $\bigcup C(\mu_i, 1)$ *containing* p *circles contains exactly* p *of the numbers* $\lambda_1, \ldots, \lambda_n$.

Then there is a permutation π *of* $\{1,\dots,n\}$ *such that for* $i=1,\dots,n$

$$|\lambda_i-\mu_{\pi(i)}|\leqslant a_n. \tag{1.8}$$

Applying this result establishes (1.6).

We finally show that the bounds given in (1.5) and (1.6) are sharp. For (1.5) consider the example

$$A=\operatorname{diag}(0,2,4,\dots,2n-2)\quad\text{and}\quad B=(2n-1)\operatorname{diag}(1,\dots,1).$$

Then $h_A(B)=1$ and $v(A,B)=2n-1$.

For (1.6) and $n=2k+1$ odd we consider

$$A=\operatorname{diag}\Big(\underbrace{0,\dots,0}_{k+1},2,4,\dots,2k\Big),$$

$$B=\operatorname{diag}\Big(1,3,\dots,2k-1,\underbrace{2k+1,\dots,2k+1}_{k+1}\Big).$$

Here $h_A(B)=h_B(A)=1$ and $v(A,B)=2k+1=n$.

If $n=2k$ even, we take

$$A=\operatorname{diag}\Big(\underbrace{0,\dots,0}_{k},2,4,\dots,2k\Big).$$

$$B=\operatorname{diag}\Big(1,3,\dots,2k-1,\underbrace{2k+1,\dots,2k+1}_{k}\Big)$$

where $h_A(B)=h_B(A)=1$ and $v(A,B)=2k-1=n-1$. ■

The importance of the bounds (1.5) and (1.6) lies in the fact that most of the bounds for $S_A(B)$ available in the literature are also bounds on $h_A(B)$ and on $\max(h_A(B),h_B(A))$ [see (2.2), (2.4), (2.6)]. Hence they also give bounds on $v(A,B)$ via (1.5) and (1.6). While those obtained by (1.5) are in the literature (see [3, 5, 11]), the bounds via (1.6) are apparently new and improve the known results by about a factor $\frac{1}{2}$. See also Remark 3 at the end of Section 4.

It should be noted that (1.5) and (1.6) stay true if $h_A(B)$ is replaced by $\max\{S_A(C(t)),\ 0\leqslant t\leqslant 1\}$, where $C(\cdot)$ is a continuous map from $[0,1]$ into the set of all complex $n\times n$ matrices such that $C(0)=A$ and $C(1)=B$.

2. BOUNDS ON THE SPECTRAL VARIATION

There are two ways for estimating the spectral variation of B with respect to A. The first and most employed uses the characteristic polynomials. If $\varphi(x)=\det(xI-A)$, $\chi(x)=\det(xI-B)$, and M is a bound on the eigenvalues of B, then it is obvious that

$$S_A(B)\leqslant\left(\max\{|\varphi(x)-\chi(x)|,|x|\leqslant M\}\right)^{1/n}. \tag{2.1}$$

This approach is used by Ostrowski in [11, Appendix K], who shows that

$$S_A(B)\leqslant(n+2)M^{1-1/n}|||A-B|||, \tag{2.2}$$

where

$$M=\max_{i,j}\left(|a_{ij}|,|b_{ij}|\right),\qquad |||A|||=\frac{1}{n}\sum_{i,j}|a_{ij}|. \tag{2.3}$$

Other bounds derived in this way are given by Bhatia and Mukherjea in [3]:

$$S_A(B)\leqslant C(n)^{1/n}M_E^{1-1/n}\|A-B\|_E^{1/n}, \tag{2.4}$$

where

$$M_E=\max(\|A\|_E,\|B\|_E),\qquad C(n)=\sum_{k=1}^{n}k^{1-k/2}\binom{n}{k}, \tag{2.5}$$

and by Bhatia and Friedland [5]:

$$S_A(B)\leqslant(2M_2)^{1-1/n}n^{1/n}\|A-B\|_2^{1/n}, \tag{2.6}$$

where

$$M_2=\max(\|A\|_2,\|B\|_2). \tag{2.7}$$

All these bounds are also bounds on $\max(h_A(B),h_B(A))$ and hence give bounds on $v(A,B)$ via (1.6), which are better by a factor $a_n/2n-1\approx\frac{1}{2}$ than those given in the abovementioned literature.

The other way of estimating $S_A(B)$ was followed by Henrici in [7]. He uses norm estimates of the resolvent $(A-\mu I)^{-1}$ and shows that

$$S_A(B) \leqslant S_n(\Delta, \|A-B\|), \tag{2.8}$$

where $\| \ \|$ is some matrix norm majorizing the spectral norm, Δ is the departure from normality of A with respect to the norm $\| \ \|$, and S_n is defined as follows: For $x \geqslant 0$ define as $g_n(x)$ the unique nonnegative root of $g+g^2+\cdots+g^n=x$. Then

$$S_n(\Delta, r)=\begin{cases} \dfrac{y}{g_n(y)}r, \quad y=\dfrac{\Delta}{r} & \text{for} \quad r>0, \\ 0 & \text{for} \quad r=0. \end{cases} \tag{2.9}$$

It is worthwhile to notice that

$$S_n(\Delta, r)=\rho(H(\Delta, r)), \tag{2.10}$$

where $H(\Delta, r)$ is the nonnegative (and for $\Delta, r>0$ also irreducible) $n \times n$ matrix

$$H(\Delta, r)=\begin{pmatrix} 0 & \Delta & & 0 \\ & \ddots & \ddots & \\ 0 & & 0 & \Delta \\ r & r & \cdots & r \end{pmatrix}. \tag{2.11}$$

From both representations it is easy to derive:

$$S_n(\Delta, r) \text{ is strictly monotone in } \Delta \text{ and } r, \tag{2.12}$$

$$r^{-1/n} S_n(\Delta, r) \text{ is strictly monotone in } \Delta \text{ and } r. \tag{2.13}$$

3. GLOBAL BOUNDS FOR $S_A(B)$ BY HENRICI'S THEOREM

The way used by Henrici is much simpler and in my opinion more adequate for bounding a geometrical magnitude like $S_A(B)$ than using the more algebraic concept of the characteristic polynomial. On the other hand the bound (2.8) involves the "local" magnitude Δ, which is not easily available. We show subsequently that this difficulty can be overcome. By using simple inequality manipulations we shall derive from (2.8) global bounds on $S_A(B)$, which are slightly better than the bounds (2.4) and (2.6), and, by the way, show that (2.8) gives normally better bounds than (2.4) and (2.6).

We start with a technical result:

LEMMA. *For given real $\tau \geqslant 0$, $\delta > 0$ and positive integer n define*

$$\gamma = \left(\delta^{n-1} + \delta^{n-2}\tau + \cdots + \tau^{n-1}\right)^{1/n}. \tag{3.1}$$

Then γ is the minimal number such that

$$\min\left(S_n(\tau M, r), \delta M\right) \leqslant \gamma M^{1-1/n} r^{1/n} \tag{3.2}$$

for all $M \geqslant 0$, $r \geqslant 0$.

Proof. Define, for fixed $M > 0$, $r_0 = \delta^n M \gamma^{-n}$. Then we have the relation

$$S_n(\tau M, r_0) = \gamma M^{1-1/n} r_0^{1/n} = \delta M. \tag{3.3}$$

Here the second equality is a consequence of the definition of r_0, while the first equality is equivalent to (3.1). From (2.13) and (3.3) we get immediately

$$r \leqslant r_0 \quad \Rightarrow \quad s(\tau M, r) \leqslant \gamma M^{1-1/n} r^{1/n} \leqslant \delta M,$$

$$r \geqslant r_0 \quad \Rightarrow \quad s(\tau M, r) \geqslant \gamma M^{1-1/n} r^{1/n} \geqslant \delta M,$$

and hence (3.2) for $M > 0$. For $M = 0$ (3.2) follows directly. (3.3) shows that γ is optimal. ■

We apply this result in several situations.

APPLICATION 1. The departure Δ from normality of A with respect to the spectral norm can be written as $\Delta = \|T\|_2$, where T is strictly upper triangular and $A = U^H(\Lambda + T)U$, U unitary, $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$, is a Schur triangular form of A. Hence

$$\Delta = \|T\|_2 \leqslant \|T + \Lambda\|_2 + \|\Lambda\|_2 \leqslant 2M_2 \qquad [\text{see } (2.7)].$$

Hence by (2.12) and (2.8)

$$S_A(B) \leqslant S_n(2M_2, \|A-B\|_2). \tag{3.4}$$

Together with the obvious inequality

$$S_A(B) \leqslant v(A,B) \leqslant 2M_2, \tag{3.5}$$

we get, applying (3.2) with $\tau=\delta=2$, $\gamma^n = n2^{n-1}$,

$$S_A(B) \leqslant \min(S_n(2M_2, \|A-B\|_2), 2M_2) \leqslant n^{1/n}(2M_2)^{1-1/n}\|A-B\|_2^{1/n}, \tag{3.6}$$

i.e. the Bhatia-Friedland bound (2.6).

APPLICATION 2. Consider the Euclidean norm $\|\ \|_E$. Ordering the eigenvalues of A and B in the following way:

$$|\lambda_1| \leqslant |\lambda_2| \leqslant \cdots \leqslant |\lambda_n|, \qquad |\mu_1| \geqslant |\mu_2| \geqslant \cdots \geqslant |\mu_n|$$

and using the inequalities $\Sigma|\lambda_i|^2 \leqslant M_E^2$, $\Sigma|\mu_i|^2 \leqslant M_E^2$ [see (2.5)], we can easily prove by induction that $|\lambda_i - \mu_i| \leqslant (1+n^{-1/2})M_E$ for $i=1,\dots,n$ and that hence

$$S_A(B) \leqslant v(A,B) \leqslant (1+n^{-1/2})M_E \tag{3.7}$$

holds. For the $\|\ \|_E$-departure Δ from normality we have

$$\Delta = \Delta_E(A) = \|T\|_E \leqslant \|\Lambda + T\|_E = M_E.$$

Hence we infer from (2.8), (2.12), (3.2) with $\tau = 1$, $\delta = 1+n^{-1/2}$, $\gamma^n = \sqrt{n}[(1+n^{-1/2})^n - 1]$,

$$S_A(B) \leqslant \min\big(S_n(M_E, \|A-B\|_E), (1+n^{-1/2})M_E\big) \tag{3.8}$$

$$\leqslant n^{1/2n}\big[(1+n^{-1/2})^n - 1\big]^{1/n} M_E^{1-1/n}\|A-B\|_E^{1/n}$$

$$\leqslant n^{1/2n}(1+n^{-1/2})M_E^{1-1/n}\|A-B\|_E^{1/n}. \tag{3.9}$$

It is not difficult to show that for all $n \geqslant 1$

$$\sqrt{n}\left[(1+n^{-1/2})^{n}-1\right]<C(n) \qquad [\text{see } (2.5)] \tag{3.10}$$

and that for all $n \geqslant 4$

$$\sqrt{n}\,(1+n^{-1/2})^{n} \leqslant C(n) \tag{3.11}$$

holds. Hence the bounds (3.8) and (3.9) improve (2.4).

Application 3. By applying Theorem 1 to the results of Applications 1, 2, global bounds for $v(A,B)$ are obtained. They can be slightly improved by applying Theorem 1 first and then the lemma. We treat only the case of the spectral norm:

As the right hand side of (3.4) or of (3.6) depends only on the bound of A, B and $\|A-B\|_2$, it is also a bound for $\max(h_A(B), h_B(A))$. Hence by applying Theorem 1 directly we get

$$v(A,B) \leqslant a_n \cdot n^{1/n}(2M_2)^{1-1/n}\|A-B\|_2^{1/n}. \tag{3.12}$$

Proceeding the other way, we first observe

$$v(A,B) \leqslant a_n S_n(2M_2, \|A-B\|_2). \tag{3.13}$$

Now using (3.5) and the lemma, we get

$$\begin{aligned} v(A,B) &\leqslant a_n \min(S_n(2M_2, \|A-B\|_2), 2/a_n M_2) \\ &\leqslant a_n(2M_2)^{1-1/n}\|A-B\|_2^{1/n}\gamma_n, \end{aligned} \tag{3.14}$$

where

$$\gamma_n = \left(1+a_n^{-1}+\cdots+a_n^{1-n}\right)^{1/n}$$

and

$$\gamma_n < \left(1+\frac{1}{n-2}\right)^{1/n} < n^{1/n} \qquad \text{for} \quad n>2. \tag{3.15}$$

This shows that (3.14) is slightly better than (3.12). A similar result is true for the Euclidean norm.

4. FINAL REMARKS

REMARK 1. The best bound on $S_A(B)$ which depends only on a common bound M_2 for $\|A\|_2, \|B\|_2$ and on the norm of the difference is obviously given by

$$S_A(B) \leqslant \partial_n M_2^{1-1/n} \|A - B\|_2^{1/n} \tag{4.1}$$

where

$$\partial_n = \sup\left\{ \frac{S_A(B)}{\|A - B\|_2^{1/n}} : \|A\|_2 \leqslant 1,\ \|B\|_2 \leqslant 1, A \neq B \right\}. \tag{4.2}$$

We have

$$2^{1-1/n} \leqslant \partial_n \leqslant n^{1/n} \cdot 2^{1-1/n} \tag{4.3}$$

where the lower bound follows from the example $A = -B =$ Identity and the upper bound is just (2.6) = (3.6). This shows that the bound (2.6) cannot be too far from the optimal bound ∂_n. No such statement can be made for the eigenvalue variation, but we suspect that the optimal bound

$$v_n = \sup\left\{ v(A, B) \|A - B\|^{-1/n} : \|A\|_2 \leqslant 1,\ \|B\|_2 \leqslant 1, A \neq B \right\}$$

is well beyond its upper bound $a_n 2^{1-1/n} n^{1/n}$ given by (3.12).

REMARK 2. Let us call bounds on $S_A(B)$ *local* if they depend explicitly on A and on some norm of $A - B$. Examples are (2.8) and the Bauer-Fike bound [1]

$$S_A(B) \leqslant \|A - B\| C(T). \tag{4.4}$$

Here, as in the sequel, $\|\ \|$ denotes a monotone vector norm and its least upper bound matrix norm $\|A\| = \sup\{\|Ax\| : \|x\| \leqslant 1\}$; furthermore, $A = T\operatorname{diag}(\lambda_i)T^{-1}$ and $C(T) = \|T\|\,\|T^{-1}\|$. The best local bound is of course given by $h(A, \|A-B\|)$, where

$$h(A,\tau)=\sup\{S_A(B):\|A-B\|\leqslant\tau\}. \tag{4.5}$$

Another problem related to this is the following. Given a number μ and a vector x, $\|x\|=1$, and $r=Ax-\mu x$, find bounds for $\min|\lambda_i-\mu|$ dependent on $\|r\|$ and on A. The best bound is $g(A,\|r\|)$, where g is given by

$$g(A,\tau)=\sup\{\min|\lambda_i-\mu|:\mathrm{glb}(A-\mu I)\leqslant\tau\} \tag{4.6}$$

and

$$\mathrm{glb}(C)=\begin{cases}\|C^{-1}\|^{-1} & \text{if } C \text{ is nonsingular,}\\ 0 & \text{otherwise.}\end{cases}$$

It is not difficult to prove that

$$g(A,\tau)=h(A,\tau). \tag{4.7}$$

In fact, if B is a matrix such that $\|A-B\|\leqslant\tau$ and μ is an eigenvalue of B, then $\mathrm{glb}(A-\mu I)\leqslant\tau$. From this we get $\min|\lambda_i-\mu|\leqslant g(A,\tau)$ and hence $h(A,\tau)\leqslant g(A,\tau)$. If on the other hand μ is such that $\mathrm{glb}(A-\mu I)\leqslant\tau$, then there exists x such that $\|x\|=1$, $\|Ax-\mu x\|=\|r\|\leqslant\tau$. Let y, x form a dual pair (see [9, p. 43]), i.e., $y^Tx=1=\|y^T\|_D\|x\|$, where $\|\ \|_D$ denotes the norm dual to $\|\ \|$. Then $B=A-ry^T$ satisfies $\|A-B\|=\|ry^T\|=\|r\|\,\|y^T\|_D\leqslant\tau$ and $Bx=\mu x$. Hence $\min|\mu-\lambda_i|\leqslant h(A,\tau)$, which implies $g(A,\tau)\leqslant h(A,\tau)$.

As a consequence of (4.7), if $f(A,\tau)$ is a function such that $S_A(B)\leqslant f(A,\|A-B\|)$ for all B, then also for all $\mu, x\neq 0$

$$\min|\lambda_i-\mu|\leqslant f\left(A,\frac{\|Ax-\mu x\|}{\|x\|}\right)$$

and vice versa. In this sense the bound (2.8) is equivalent to the bound of Morrison [10] (see also [6]), and the bound (4.4) is equivalent to the result

$$\min|\lambda_i-\mu|\leqslant C(T)\frac{\|Ax-\mu x\|}{\|x\|}$$

in [2].

REMARK 3. We finally remark that Theorem 1 leads also to an improvement of Ostrowski's theorem [11, Appendix A] on the continuity of the roots of algebraic equations:

Let $f(x)=\sum_{v=0}^{r}a_{v}x^{n-v}$, $g(x)=\sum_{v=0}^{n}b_{n}x^{n-v}$, $a_0=b_0=1$,

$$\gamma=2\max_{v>0}\left(|a_v|^{1/v},|b_v|^{1/v}\right)\quad and \quad \varepsilon=\left(\sum_{v=1}^{n}|a_v-b_v|\gamma^{n-v}\right)^{1/n}.$$

Then the roots $x_1,\dots,x_n$ *of* f *and* $y_1,\dots,y_n$ *of* g *can be ordered so that for* $i=1,\dots,n$,

$$|x_i-y_i|\leqslant\varepsilon\cdot\begin{cases} n, & n\ odd,\\ n-1, & n\ even.\end{cases}$$

Similarly the results in [4] can be improved.

REFERENCES

1 F. L. Bauer and C. T. Fike, Norms and exclusion theorems, *Numer. Math.* 2:137–141 (1960).

2 F. L. Bauer and A. S. Householder, Moments and characteristic roots, *Numer. Math.* 2:12–53 (1960).

3 R. Bhatia and K. Mukherjea, On the rate of change of spectra of operators, *Linear Algebra Appl.* 27:147–157 (1979).

4 R. Bhatia, On the rate of change of spectra of operators. II, *Linear Algebra Appl.*, to appear.

5 R. Bhatia and S. Friedland, Variation of Grassman powers and spectra, *Linear Algebra Appl.*, 40:1–18 (1981).

6 L. Elsner, Einschliessungssätze für Eigenwerte nichtnormaler Matrizen, in *ISNM* 7, Birkhäuser, Basel, 1967, pp. 185–193.

7 L. Elsner, C. Johnson, J. Ross, and J. Schönheim, A generalized matching problem arising in estimating the eigenvalue variation of two matrices, submitted for publication.

8 P. Henrici, Bounds for iterates, inverses, spectral variation and fields of values of non-normal matrices, *Numer. Math.*, 4:24–40 (1962).

9 A. S. Householder, *The Theory of Matrices in Numerical Analysis*, Blaisdell, 1964.

10 D. Morrison, Errors in the solution of eigenvalue problems by finite difference methods, Ph.D. Dissertation, Univ. of California, Los Angeles, 1961.

11 A. Ostrowski, *Solution of Equations in Euclidean and Banach Spaces*, Academic, 1973.

Received September 1981